

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/721,479B

DATE: 05/03/2001
TIME: 13:19:34

Input Set : A:\seqlist.txt
Output Set: N:\CRF3\05032001\I721479B.raw

ENTERED

3 <110> APPLICANT: Coit, Doris
4 Medina-Selby, Angelica
5 Selby, Mark
6 Houghton, Michael
8 <120> TITLE OF INVENTION: NOVEL HCV NON-STRUCTURAL POLYPEPTIDE
10 <130> FILE REFERENCE: PP01617.002
12 <140> CURRENT APPLICATION NUMBER: US 09/721,479B
13 <141> CURRENT FILING DATE: 2000-11-22
15 <160> NUMBER OF SEQ ID NOS: 19
17 <170> SOFTWARE: PatentIn Ver. 2.0
19 <210> SEQ ID NO: 1
20 <211> LENGTH: 9620
21 <212> TYPE: DNA
22 <213> ORGANISM: Artificial Sequence
24 <220> FEATURE:
25 <221> NAME/KEY: CDS
26 <222> LOCATION: (1990)..(7302)
28 <220> FEATURE:
29 <223> OTHER INFORMATION: Description of Artificial Sequence: Hepatitis C pns345
31 <400> SEQUENCE: 1
32 cgcgcgctttc ggtgatgacg gtgaaacact ctgacacatg cagctcccgg agacgggtcac 60
34 agcttgctctg taagcggatg cggggagcag acaagcccgt cagggcgctg cagcgggtgt 120
36 tggcgggtgt cggggctggc ttaactatgc ggcacacagc cagattgtac tgagagtgc 180
38 ccatatgaag ctttttgcaa aagcctaggc ctccaaaaaa gcctcctcac tacttctgga 240
40 atagctcaga ggccgaggcg gcctcggcct ctgcataaat aaaaaaaatt agtcagccat 300
42 ggggcggaga atgggcggaa ctgggcgggg aggggaattat tggctattgg ccattgcata 360
44 cgttgtatct atatcataat atgtacattt atattggctc atgtccaata tgaccgccat 420
46 gttgacattg attattgact agttattaat agtaatacat tacgggggtca ttagttcata 480
48 gccatatat ggagttccgc gttacataac ttacggtaaa tggcccgctt ggctgaccgc 540
50 ccaacgaccc cggcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 600
52 ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagtac 660
54 atcaagtgtg tcatatgcca agtccgcccc ctattgacgt caatgacggt aaatggccc 720
56 cctggcatta tgcccagtac atgaccttac gggactttcc tacttggcag tacatctacg 780
58 tattagtcac cgctattacc atggtgatgc ggttttgcca gtacaccaat gggcgtggat 840
60 agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt 900
62 tttggcaca aaatcaacgg gactttccaa aatgtcgtaa taaccccgcc cgttgacgc 960
64 aaatgggcgg taggcgtgta cgggtggagg tctatataag cagagctcgt ttagtgaacc 1020
66 gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga caccgggacc 1080
68 gatccagcct cgcgcggcgg gaacggtgca ttggaacgcg gattccccgt gccagagt 1140
70 acgtaagtac cgcctataga ctctataggg acaccccttt ggctcttatg catgctatac 1200
72 tgtttttggc ttggggccta tacaccccg ctccttatgc tatagggtgat ggtatagctt 1260
74 agcctatagg tgtgggttat tgaccattat tgaccactcc cctattgggt acgatacttt 1320
76 ccattactaa tccataacat ggctctttgc cacaactatc tctattggct atatgccaat 1380
78 actctgtcct tcagagactg acacggactc tgtattttta caggatgggg tccatttatt 1440
80 atttacaat tcacatatac aacaacgcgg tcccccggtc ccgcagtttt tattaaacat 1500
82 agcgtgggat ctccgacatc tcgggtacgt gttccggaca tgggctcttc tccggtagcg 1560
84 gggagcttc cacatccgag ccctgggtccc atccgtccag cggctcatgg tcgctcgga 1620

RAW SEQUENCE LISTING

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,479B

TIME: 13:19:34

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\05032001\I721479B.raw

```

86 gtccttgcct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680
88 gtgtgccgca caaggccgtg gcggtagggg atgtgtctga aaatgagctc ggagattggg 1740
90 ctgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800
92 agttgttgta ttctgataag agtcagaggt aactcccggt gcggtgctgt taacgggtgga 1860
94 gggcagtgta gtctgagcag tactcgttgc tgccgcgcgc gccaccagac ataatagctg 1920
96 acagactaac agactgttcc ttcccatggg tcttttctgc agtcaccgtc gtcgacctaa 1980
98 gaattcacc atg gct gca tat gca gct cag ggc tat aag gtg cta gta ctc 2031
99      Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu
100      1          5          10
102 aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac atg tcc aag 2079
103 Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys
104 15          20          25          30
106 gct cat ggg atc gat cct aac atc agg acc ggg gtg aga aca att acc 2127
107 Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr
108      35          40          45
110 act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc ctt gcc gac 2175
111 Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp
112      50          55          60
114 ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt gac gag tgc 2223
115 Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys
116      65          70          75
118 cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act gtc ctt gac 2271
119 His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp
120      80          85          90
122 caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc acc gcc acc 2319
123 Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr
124 95          100          105          110
126 cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag gag gtt gct 2367
127 Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala
128      115          120          125
130 ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct atc ccc ctc 2415
131 Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu
132      130          135          140
134 gaa gta atc aag ggg ggg aga cat ctc atc ttc tgt cat tca aag aag 2463
135 Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys
136      145          150          155
138 aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc atc aat gcc 2511
139 Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala
140      160          165          170
142 gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccg acc agc ggc 2559
143 Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly
144 175          180          185          190
146 gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc tat acc ggc 2607
147 Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly
148      195          200          205
150 gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc cag aca gtc 2655
151 Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val
152      210          215          220
154 gat ttc agc ctt gac cct acc ttc acc att gag aca atc acg ctc ccc 2703

```

RAW SEQUENCE LISTING

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,479B

TIME: 13:19:34

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\05032001\I721479B.raw

```

155 Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro
156          225          230          235
158 caa gat gct gtc tcc cgc act caa cgt cgg ggc agg act ggc agg ggg 2751
159 Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly
160      240          245          250
162 aag cca ggc atc tac aga ttt gtg gca ccg ggg gag cgc ccc tcc ggc 2799
163 Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly
164 255          260          265          270
166 atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca ggc tgt gct 2847
167 Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala
168          275          280          285
170 tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta cga gcg tac 2895
171 Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr
172          290          295          300
174 atg aac acc ccg ggg ctt ccc gtg tgc gag cag cat ctt gaa ttt tgg 2943
175 Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp
176          305          310          315
178 gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac ttt cta tcc 2991
179 Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser
180      320          325          330
182 cag aca aag cag agt ggg gag aac ctt cct tac ctg gta gcg tac caa 3039
183 Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln
184 335          340          345          350
186 gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg tgg gac cag 3087
187 Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln
188          355          360          365
190 atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat ggg cca aca 3135
191 Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr
192          370          375          380
194 ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc acc ctg acg 3183
195 Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr
196          385          390          395
198 cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc gac ctg gag 3231
199 His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu
200      400          405          410
202 gtc gtc acg agc acc tgg gtg ctc gtt gcc ggc gtc ctg gct gct ttg 3279
203 Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu
204 415          420          425          430
206 gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg ggc agg gtc 3327
207 Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val
208          435          440          445
210 gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa gtc ctc tac 3375
211 Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr
212          450          455          460
214 cga gag ttc gat gag atg gaa gag tgc tct cag cac tta ccg tac atc 3423
215 Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile
216      465          470          475
218 gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag gcc ctc ggc 3471
219 Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly

```

RAW SEQUENCE LISTING

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,479B

TIME: 13:19:34

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\05032001\I721479B.raw

220	480	485	490	
222	ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc cct gct gtc	3519		
223	Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val			
224	495 500 505 510			
226	cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag cat atg tgg	3567		
227	Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp			
228	515 520 525			
230	aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca acg ctg cct	3615		
231	Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro			
232	530 535 540			
234	ggg aac ccc gcc att gct tca ttg atg gct ttt aca gct gct gtc acc	3663		
235	Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr			
236	545 550 555			
238	agc cca cta acc act agc caa acc ctc ctc ttc aac ata ttg ggg ggg	3711		
239	Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly			
240	560 565 570			
242	tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct act gcc ttt gtg	3759		
243	Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val			
244	575 580 585 590			
246	ggc gct ggc tta gct ggc gcc gcc atc ggc agt gtt gga ctg ggg aag	3807		
247	Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys			
248	595 600 605			
250	gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg gcg gga gct	3855		
251	Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala			
252	610 615 620			
254	ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc acg gag gac	3903		
255	Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp			
256	625 630 635			
258	ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc ctc gta gtc	3951		
259	Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val			
260	640 645 650			
262	ggc gtg gtc tgt gca gca ata ctg cgc cgg cac gtt ggc ccg ggc gag	3999		
263	Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu			
264	655 660 665 670			
266	ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc tcc cgg ggg	4047		
267	Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly			
268	675 680 685			
270	aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat gca gct gcc	4095		
271	Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala			
272	690 695 700			
274	cgc gtc act gcc ata ctc agc agc ctc act gta acc cag ctc ctg agg	4143		
275	Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg			
276	705 710 715			
278	cga ctg cac cag tgg ata agc tcg gag tgt acc act cca tgc tcc ggt	4191		
279	Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly			
280	720 725 730			
282	tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg ttg agc gac	4239		
283	Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp			
284	735 740 745 750			

RAW SEQUENCE LISTING

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,479B

TIME: 13:19:34

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\05032001\I721479B.raw

```

286 ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg cct ggg atc 4287
287 Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile
288 755 760 765
290 ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg cga ggg gac 4335
291 Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp
292 770 775 780
294 ggc atc atg cac act cgc tgc cac tgt gga gct gag atc act gga cat 4383
295 Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His
296 785 790 795
298 gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc tgc agg aac 4431
299 Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn
300 800 805 810
302 atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg ggc ccc tgt 4479
303 Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys
304 815 820 825 830
306 acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg agg gtg tct 4527
307 Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser
308 835 840 845
310 gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc cac tac gtg 4575
311 Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val
312 850 855 860
314 acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag gtc cca tcg 4623
315 Thr Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser
316 865 870 875
318 ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat agg ttt gcg 4671
319 Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala
320 880 885 890
322 ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc aga gta gga 4719
323 Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly
324 895 900 905 910
326 ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag ccc gaa ccg 4767
327 Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro
328 915 920 925
330 gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc cat ata aca 4815
331 Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr
332 930 935 940
334 gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc ccc tct gtg 4863
335 Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val
336 945 950 955
338 gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc aag gca act 4911
339 Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr
340 960 965 970
342 tgc acc gct aac cat gac tcc cct gat gct gag ctc ata gag gcc aac 4959
343 Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn
344 975 980 985 990
346 ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg gtt gag tca 5007
347 Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser
348 995 1000 1005
350 gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt gtg gcg gag 5055

```

VERIFICATION SUMMARY

DATE: 05/03/2001

PATENT APPLICATION: US/09/721,479B

TIME: 13:19:35

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\05032001\I721479B.raw